Development of Law Enforcement System for Detection of Driver's Seatbelt and Speed Estimation through Machine Vision

Author

Syed Muhammad Ali Haider

171608

Supervisor

Dr. Waqar Shahid Qureshi

DEPARTMENT OF MECHATRONICS ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

JULY, 2020

Development of Law Enforcement System for Detection of Driver's Seatbelt and Speed Estimation through Machine Vision

Author

Syed Muhammad Ali Haider

171608

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Mechatronics Engineering

Thesis Supervisor:

Dr. Waqar Shahid Qureshi

Thesis Supervisor's Signature:_______________________________________

DEPARTMENT OF MECHATRONICS ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

JULY, 2020

Declaration

I certify that this research work titled "*Development of Law Enforcement System for Detection of Driver's Seatbelt and Speed Estimation through Machine Vision"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged / referred.

Signature of Student

Syed Muhammad Ali Haider

MS-MTS-171608

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Syed Muhammad Ali Haider

MS-MTS-171608

Signature of Supervisor

Copyright Statement

Acknowledgements

I am thankful to Allah to have guided me throughout this work at every step and for every new thought. Indeed I could have done nothing without His help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am thankful to my parents who raised me and continued to support me throughout in every walk of my life.

I would also like to express special thanks to my supervisor, Dr. Waqar Shahid, for his generous help and guidance throughout my thesis.

I would also like to thank Dr. Umer Shahbaz, and Dr. Mohsin Islam Tiwana for being in my thesis guidance and evaluation committee and express my special thanks to my younger brother Syed Muhammad Ali Qasim for facilitating me wherever I required his help. I am also thankful to my cousin, Syed Ali Abid for his support and cooperation in video recording and being there with me whenever I required him.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional parents, family members and generous teachers whose tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Due the rapid growth in number of vehicles in Pakistan, traffic analysis has become a need of hour. Rate of accidents has gradually increased due to over-saturated roads, inexperienced drivers, unplanned movement and lack of respect to road safety rules. Vehicle safety condition is quite bad and necessities like airbags and reliable braking system are still considered as “luxuries” by vehicle manufacturers. Detection of vehicle speed and driver’s seatbelt, for safety concerns, have therefore become an essential requirement for traffic law enforcement agencies. This report presents law enforcement system for detection of vehicle speed and driver’s seatbelt based on machine vision. HAAR Cascade Classification is used for speed estimation and seatbelt detection. Vehicle tracking along the path in which vehicle enters the frame till the point when vehicle leaves the frame is required for speed estimation. Based on training images of our own data set collected at various locations, algorithm looks for possible seatbelt area in the upper body of driver’s area of vehicle and saves the detected image in a folder with pre-defined path. Assuming that vehicles are symmetric in shape, vehicle path is flat and monocular camera has negligible distortion, detection is done considering driving parameters in Pakistan. It has been found in results that detection, whether driver’s seatbelt is ON or not is about 87 and speed estimation is accurate by 90 percent.

Key Words: *Seatbelt, Speed, Monocular camera, HAAR Cascade*

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

Road accident is regarded as an unfortunate event with no proper cause. Most of the accidents occur due to unprovoked rashness on daily basis. These are not proper accidents and can be minimized with care and attentiveness at driver's part. It is often advised by Traffic and Road Safety Department that ''*Speed Thrills but Kills''* but general public turns deaf ear towards it. Traffic safety consists of 3E's: **education, enforcement** and **engineering**. Road safety studies carried out in previous years indicate that road accidents have increased in under-developed and developing countries and decreased in developed countries.

Driving characteristics of a driver is also an actor in traffic accidents. Some mistakes happening at the driver's end might appear as a cause of crash like in-attentiveness due to phone calls, drowsiness due to fatigue, rash driving, driving on wrong side of the road, inappropriate or no use of indicator lights and listening music while driving. Such careless approach reduces the chances of taking correct decisions at the right time. Equipment breakdown like tire bursts, suspension or steering breakout or brake fails may also cause road side accidents. WHO report on road safety estimated that approximately 1.2 million are killed and 50 million got injured each year in road accidents. Projections approximate 65% increase in forth coming years. [1]

Education and Enforcement are as important in road safety measures as Engineering although education and enforcement are unsatisfactory compared to engineering. There is lack of organized setup for education and enforcement of road safety rules and regulations. Unfortunately in Pakistan, we have little traffic sense and literally meagre respect for road safety regulations. Statistics reveal that there lies a close relationship between driver's instinct and type of accident. Studies carried out in recent years revealed that in Pakistan, Punjab faces highest and Baluchistan suffers least number of road accidents moreover Sindh has highest number of fatal accidents. [1] Bad road conditions and fast driving are the major causes of traffic collisions but the situation is expected to get worse with every passing day if appropriate attention is not paid to the issue. The rise in number of vehicles and saturation of roads is also a cause of road side accidents. Lack of planned movement and saturated roads cause an urge among drivers to violate traffic regulations. Statistics reveal that by following the speed limits, the probability of occurrence of accidents is greatly reduced. Nilsson's "Power Model'' shows that increase in 1% average speed increases crash occurrence frequency by 2% , increase in injury crash frequency by 3% and crash frequency of fatal accidents by 4%. Decreasing speed by a few km/h subsequently reduces risk and severity of collisions and ensures safety of life. [2]

It has been seen that drivers somehow might follow maximum speed limits but seatbelt is paid the least respect in Pakistan. A significant majority of drivers do not bother wearing seatbelt despite knowing its importance. For a driver seatbelt has a great significance. Among front-seat passengers, seat belt lessens the death probability by 45%, cuts the serious injury risk by 50%. Seatbelts work on the principles of physics, when a vehicle is travelling at a speed 'A km/h', the body also travels at the same speed 'A km/h'. Accident may stop the travelling vehicle but body keeps moving at the same speed, which may propel the body fly out of the vehicle at speed 'A km/h'. Seatbelt holds the body in place and increases the time of impulse. Seatbelt distributes momentum and impulse into chest, shoulders and pelvis which may leave the body bruised but serious injuries are avoided. Seat belt prevents ejection of passengers out of the vehicle during a crash. Individual not wearing seat belt increases his risk of being ejected out from the vehicle by 30 during a crash. Seatbelts have saved an approximate of 300,000 lives since 1975. [3] [4]

Keeping in view the current traffic situation and increased accident cases in Pakistan, we are interested in developing an efficient and cost effective device that can differentiate between a driver with seatbelt or without a seatbelt and is capable of detection whether the driver is driving within the safe speed limit or is exceeding the limits. Speed cameras and Seatbelt detecting cameras are available but a single integrated device capable of detecting both seatbelt and speed is not commonly seen. We are looking to fulfill this gap through our research.

1.1 Objectives of Research

In this research we are interested in finding out:

1. Detecting if the driver is wearing a seat belt, from a video feed of a monocular camera.
2. Detection of vehicle speed from a monocular camera video feed.
3. Determining mounting angle of monocular camera for appropriate results.

1.2 **Proposed Solution**

Major methods of seatbelt detection were studied. We intend to reach our goal through Haar Cascade Classification. A brief summary of the solution is as follows:

1. To develop Haar Cascade Classifier for seatbelt and vehicle detection.
2. Training Haar Cascade.
3. Apply bounding box on each four tire vehicle on the road.
4. Track the path of each vehicle from the point “A” when vehicle enters the frame till the point “B” when vehicle leaves the frame, in a fixed distance.
5. Detect seatbelt in the video-feed.
6. Differentiate distance with respect to time for finding speed of vehicle.

The research will take into account the method of developing an appropriate classifier and determining required path to achieve our goal.

2. LITERATURE REVIEW for SEATBELT and SPEED DETECTION

Approximately 1.2 million people face causalities each year due to road side accidents. Seatbelt is a protective measure and probability of serious incidents decreases manifolds by using seatbelts. In the past four to five years much work is done on autonomous seatbelt detection through image processing and computer vision. This section of report presents a glance of previous efforts done by various researchers.

Jing Xu, et al devised an improvement in Adaptive Boosting (Adaboost) algorithm to detect line segment as a part of seatbelt using LSD. They used template matching to see a seatbelt patch in the image using sliding window approach. Input image was compared with the template image and results were saved in a matrix, pixel brightness represented degree of matching with template image. [5]

Huiwen Guo, et al devised a method of detecting seatbelt in images. First driver's area in the vehicle was isolated from the image. Second step was to determine the seatbelt edges after getting the information from HSV color space. Further verification of edges was also done to get more accurate results. The first part was to convert the image into HSV color space. Seatbelt lies at an angle between 50° to 70° according to standards. Using Hough transform, a line more than 100 pixels long with allowed discontinuity of not more than 20 pixels, two longest parallel lines were searched in the image present at an angle lying between 50° to 70°. The indicated region would be a seatbelt. [6]

Benjamin Penchas, et al used deep learning to estimate vehicle speed using their own optical flow object detection model using vehicle mounted camera. Pre-training on ImageNet was done considering the importance of scale. Convolutional layers use 64 by 3 filters with stride of two and ReLU activation layers. The final fully connected layer had 15 neurons and no activation function. Softmax function was applied to final layer to get discrete speeds. [7]

Zhongjian, et al used LC-RNN deep learning model for vehicle speed estimation. They integrated both CNN and RNN to look for the speed variations with reference to surrounding areas. Detection accuracy was improved by fusing factors like periodicity and context factors. The model however failed to consider topological structure of road network. [8]

Witold Czajewski, et al used image processing to determine the speed of approaching vehicle. The method was to measure the distance covered by a vehicle between two frames. The camera may or may not be triggered by a radar to take pictures from a specific distance. Pictures can be taken by a mounted camera and detection could be done in a continuous video frame. [9]

Dogan, et al estimated vehicular speed from side-view images taken by an un-calibrated camera. They captured consecutive video frames and found ROI in the images, background subtraction was done through histogram thresholding. By tracking foreground points in the image, they estimated velocity vectors of various reference points on the vehicle from frame images. Displacement vectors of reference points were computed with respect to time to get velocity vectors. By measuring mean and standard deviation in the velocity vectors, average speed of vehicle was calculated. [10]

Azhar Hussain, et al determined speed of multiple vehicles in real time using fuzzy logic. In the first step, they extracted background in specific frames by using adaptive median filter. They extracted foreground using fuzzy logic by subtraction of background and current frame pixels. If background pixels matched with foreground pixels, fuzzy rule set was defined. Defuzzification was performed using α-cut method. They determined vehicle speed by using MAP (maximum posterior probability) estimator. [11]

Wimalaratna and Sonnadara estimated speed of mobile vehicles from video frame. Pre-processing was done in gray scale and median filters were used. Sobel edge detector was applied and consecutive image differencing was used for vehicle extraction. Pixel difference in two images gave a new image. Noise in the image was removed using intersection of two edged images. Erosion and Dilation was applied to get the foreground. Aforge.net library in MATLAB was used to get vehicle positioning. Speed of vehicle was calculated using vehicle tracking. [12]

Bhagyashri and Pravesh determined vehicle speed in video sequences. Moving pixels in video frame were detected in binary form using background subtraction and inter-frame differencing. Vehicle tracking was done to compute distance travelled by vehicle in consecutive frames. Acquired distance was divided with time taken to get vehicle speed. [13]

Dolley and Ekta estimated vehicle’s speed by tracking its motion in video frames using Lukas Kanade algorithm. Motion estimation was done by matching the regional intensities of pixels in consecutive frames. Gradient based edge detection was done to calculate image flow. Optical flow model was used to get velocity vectors in matrix form, region filtering and thresholding was done on the matrix. Camera calibration was done to convert 2D coordinates to 3D world coordinates. Eucledian distance between two successive matrix elements was calculated. This distance was converted to actual distance using pixel to distance ration. Dividing number of frames by frames per second gives actual time. Distance time relation was used to compute vehicle speed. [14]

Tingting Huang used Faster R-CNN with ResNet 101 with previously trained COCO model for multiple object detection in a video. He used tracking algorithm based upon histogram comparison and finally used wrapping method to find vehicle speed from pixels to the physical world. Histogram based tracker connects frames by finding minimum Chi-squared distance among a group of candidates. Chi-squared distance gives high variation and distinguishability and is easy to implement. Wrapping with linear perspective transformation converted pixel per second to mile per hour for speed calculation. [15]

Wenchang, et al estimated vehicle speed using monocular camera. They did detection and classification of vehicle of interest. Based on template matching, they did object tracking to compute frame to frame distance. Frame to frame distance was converted to physical distance by using pixel to distance ration. This was done by projective transformation through manual camera calibration. By dividing distance over time they determined vehicle speed. [16]

3. SPEED ESTIMATION DEVICES

Several commercial portable speed measurement devices are available in the market including intrusive, non-intrusive and off-road devices.

- Intrusive devices include pizeo-electric and pneumatic tubes, magnetic detectors and bending plates.
- Non-intrusive devices include infrared or ultrasonic sensors, MW detectors or video cameras.
- Off-road devices include probe vehicle, LASER or Lidar guns.

These devices have specific advantages and dis-advantages. These devices are either placed on the road surface or on the road sides or hand held. Certain discrepancies in measurements are possible as well. Given below is a brief survey of speed measurement devices.

3.1.1 Automated Traffic Classifiers (ATC)

i. At a specific site on the road, two road tubes are installed.
ii. Tubes are connected to sensors and sensors are in contact with the micro-processor.
iii. Multiple types of sensor data is processed and stored by the micro-processor. [11]

Data is retrieved by using a computer. The figure below shows a road tube.

Figure 1: Inductive Loop [11]

3.1.2 LIDAR and Radar Gun

i. LIDAR and Radar guns are included in manual devices with an operating range lying between 60 to 300 meters.
ii. They require line of sight and a person for proper operation.
iii. They can operate well in both clear and bad weather. They use infrared pulses to penetrate glass and find speed and range of target. [11]

The figure below LIDAR and Radar guns.

Figure 2: LIDAR Gun [11]

3.1.3 Manual Counter

i. Run speed is found by using distance time relation. By dividing distance by time we get speed.
ii. Trap length is the patch on a road where time is calculated for a vehicle to cross. This patch is also called test section.
iii. An observer measures the time when a vehicle enters the patch and stops when the vehicle leaves the trap length. [11]

Figure below shows a manual counter.

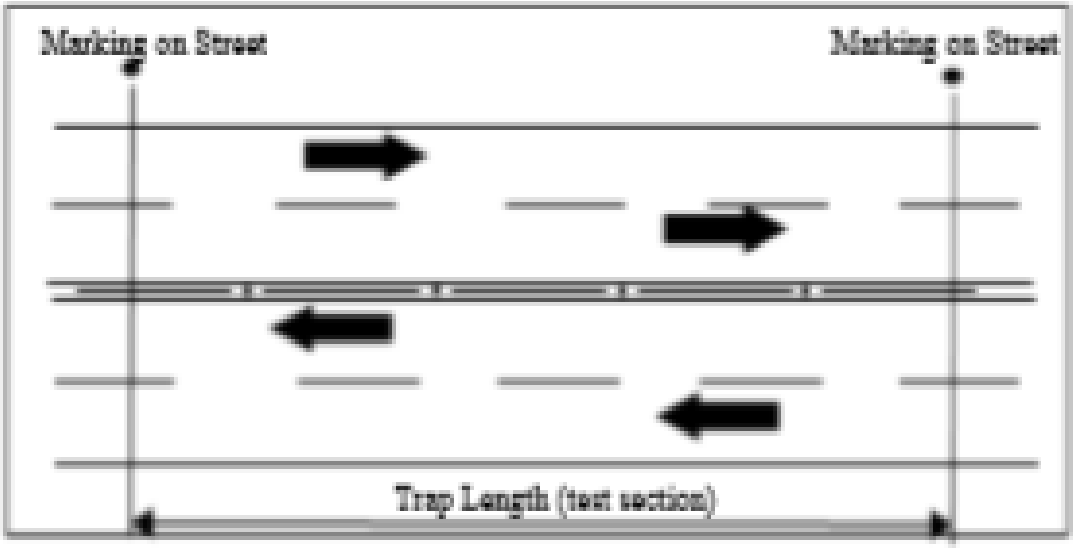

Figure 3: Manual Counter [11]

3.1.4 Probe Vehicle

i. PV technologies are a part of intelligent transportation system. They are used for real time monitoring and also for route guidance. Real time data is also collected by them.
ii. Probe vehicles have GPS transmitter and receiver system.
iii. Space, control and user are the components of a typical GPS system.
iv. They pick up and send signals including speed and route to orbiting satellites. [11]

The figure below shows a probe vehicle data management system.

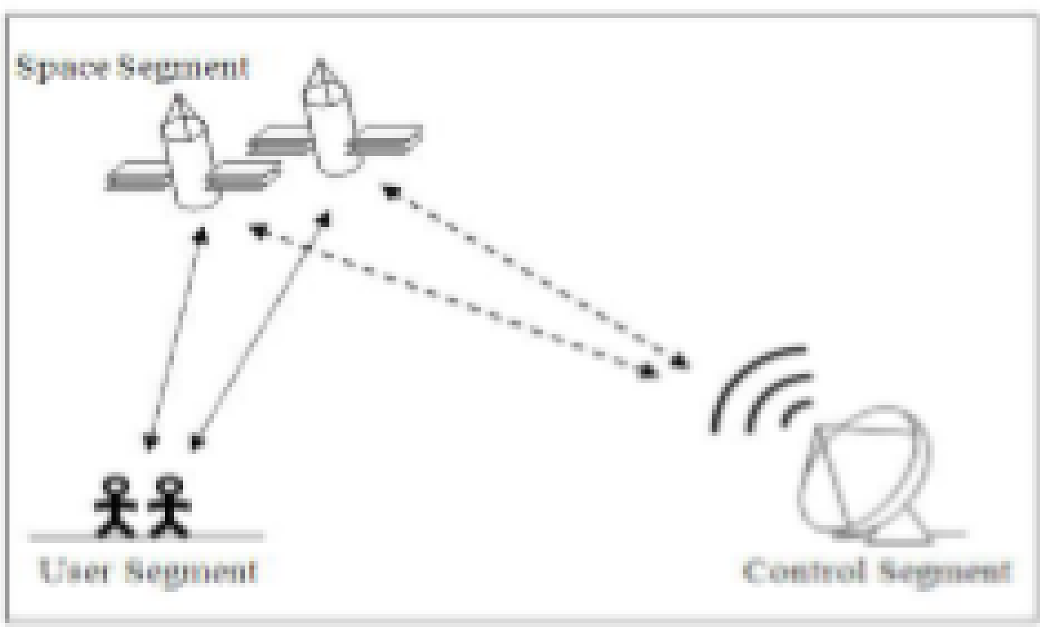

Figure 4: Probe Vehicles [11]

4. HAAR CASCADE CLASSIFICATION

Object Detection is a feature related to computer vision, deep learning, image processing and deals with object detection in images and videos at a particular instant of time. Haar Cascade Classifiers effectively works for object detection. Machine learning algorithm called ***Haar Cascade,*** *in which a set of positive and negative images are used to train a cascade function.* It is used for object detection in other video sequences and images afterwards. The method was proposed by Viola and Jones. Positive and Negative Images are required to train a classifier of required object detection.

- **Positive images –** contain objects which are required to be identified.
- **Negative Images –** a set of all images which are not required to be detected. [17]

Haar Cascade can detect any object as long as we have XML file required for appropriate detection. The algorithm comprises of four stages:

1. Feature Selection
2. Integral Images Creation
3. Training Adaboost
4. Cascading Classifiers
    - First step in cascade designing is the Haar Features collection.
    - Haar feature emphasizes rectangular regions in a specific location in one detection window, pixel intensities in each region are summed up and difference of sums is calculated.
    - Integral Images make the calculation super-fast.
    - Many of the features are irrelevant. Best out of all features are selected.
    - It is done by using Adaptive Boosting (Adaboost) Algorithm.
    - Adabost Algorithm selects best suited features and trains the classifier upon them. [17]
    - Assigns weights to training dataset.
    - Calculates variance of all samples.
    - Selects a weak classifier and goes constantly from weak to strong classifier.
    - For all positive and negative samples, computes eigenvalues and weighted error rates.
    - Classifier with smallest weighted error is chosen as a weak classifier.
    - Attains a strong classifier after repeated iterations.

5. OBJECT DETECTION

Object detection estimates the locations of different objects and classifies them as they are, based on features like edges, corners and lines across images and combines them to form complex shapes. Due to availability of different datasets and pre-trained neural networks, it is quite possible to achieve object detection through deep neural networks without a specialized hardware. Preprocessing data into simpler variables space is known as feature extraction. Detectors extract features of data, and are used as inputs to the learning algorithms. [18]

5.1 Types

Supervised learning is a general way of using machine learning. In this type of object detection, training images are used with marked classes and locations of objects to be detected. The algorithm is capable of predicting the labels of unseen data after learning. Classification deals with predicting correct classes of testing data on the basis of training data. [18]

In **Unsupervised learning**, useful properties of the data are learnt by the algorithm without a supervisor telling about the correct output, for example clustering. Now a days, unsupervised preprocessing is a tool for discovering meaningful representation of data. [18]

5.2 Training

i. To train a neural network, weights of neurons are selected to achieve target outputs from learnt inputs. **Back-propagation algorithm** provides an efficient solution in calculating weights iteratively. The **gradient descent** is a common optimization method.

ii. In **first phase** of algorithm, input vector is propagated through neural network.

iii. The output of network is compared with the desired output using any loss function.

iv. The gradient of the loss function called **error value** is then computed. The error values are propagated through network for calculating error values of hidden layer's neurons.

v. The loss function gradients of hidden neurons are computed using chain rule of differentiation. Finally weights of the neurons are updated. This is called the **learning rate**.

vi. After updated weights are picked up by the algorithm and continues execution with a different input till the weights converge.

- The explained method is called **online learning** which computes updates in weights after each updated input. [18]

- Other way of getting updates is called **full batch learning**, in which we obtain weight updates in dataset. This is computationally quite heavy. [18]

- **Mini-batch learning** is a compromised version is, in which we only use a portion of training dataset for every update. [18]

5.3 Object Detection

- Object detection involves developing a solution invariant to deformations, changes in light intensities as well as classifying and locating regions in an image. We need to perform image segmentation to have possible location of objects in a scene. [18]
- For locating an object in a frame, we need to know its shape.
- The size and location is defined using a bounding box.
- The sub-images located in bounding box are classified by algorithm later on, trained using machine learning.
- Object boundaries are iteratively refined, by making initial guesses.
- Popular feature detectors before CNNs were SIFT and HOG.
- There were two solutions for obtaining bounding boxes.

i. A region proposals set is generated and most of the region proposals are rejected later on, in the first solution. This involves using sliding window detector. [18]

ii. We generate a sparse bounding boxes set in second method, like in Selective Search method.

- Combining sparse regions with CNNs also provides good results. [18]

5.4 Selective Search Algorithm

- Selective Search algorithm uses sparse object location set of an image using hierarchical partitioning.
- The algorithm uses Graph Based Image Segmentation by creating a set of super-pixels. Selective Search continues by grouping together the regions iteratively, starting with two most similar regions.
- Many measures compute the similarity index. These measures include color similarity (computing color histogram), texture similarity (computing SIFT), region size (small regions are merged earlier) etc.
- The object locations generated are arranged by likelihood of location, containing an object.
- To prevent large objects from being favored too much, an element of randomness is added. Lower-ranking duplicates are removed.
- Complementary color spaces and complementary starting regions are used ensure lightning invariance. [18]

5.5 Edge Boxes

- The number of edge contours enclosed by a bounding box is correlated with possibility of an object present in it.
- The edge map is drawn using Structured Edge Detector.
- Thick edge lines are made thin by using Non-Maximum Suppression.
- Scanning the image gives region proposals by using sliding window method and outputs object presence score and scale.
- The summation of edge strengths lying completely in a bounding box and subtracting the edge strengths crossing the boundary gives score of presence of an object.
- High score regions are further refined later on. [18]

6. CAMERA CALIBRATION

Camera calibration and camera modelling are two major tasks of computer vision. A camera model relates 3D object space to 2D image space. The procedure used to calculate camera parameters is called camera calibration. The procedure uses several images for reference points. Regardless of camera model being used, camera parameters need to be calculated. Usually calculations are based on the reference point images. [19]

The position of 3D reference points may or may not be known in advance. With unknown initial reference positions, the procedure is **auto-calibration**. **Bundle adjustment** procedure calculates camera parameters and 3D coordinates of reference points from multiple images. **True auto-calibration method** does not require calculation of reference points, however uses just the information of image points corresponding to same 3D points in multiple images. In the true auto-calibration, many methods of projective theory (absolute conic, dual absolute conic, plane at infinity) are used. Calibration method that requires measurement information of poses or positions of the camera is called **active calibration**. In active calibration, hand-eye relationship along with reference positions and camera parameters is determined. [19]

Camera model is used to project 3D points as reference to the image. By summation of square of difference of the detected and calculated image points is used as an error criterion. To find the correspondence between points in object space and image space, feature points are extracted using SIFT (scale invariant feature transform) or SURF (speed up robust features) algorithm. Checker-board pattern can also be used to know, which reference is which. [19]

As center of gravities of 3D reference points do not correspond to center of gravities of corresponding 2D points, pre-processing is required.

- Initially, 2D image references around 2D image co-ordinates are extracted.
- Camera parameters along with 3D reference positions, are found using optimization method, which match 2D points correspond to object lines.
- Good match gives corresponding parameters as a start data.
- Back projection of 3D reference centers using appropriate camera model gives image co-ordinates.
- Forward and reverse camera model projections are required in these procedures. [19]

Block diagram ahead, shows camera calibration procedure:

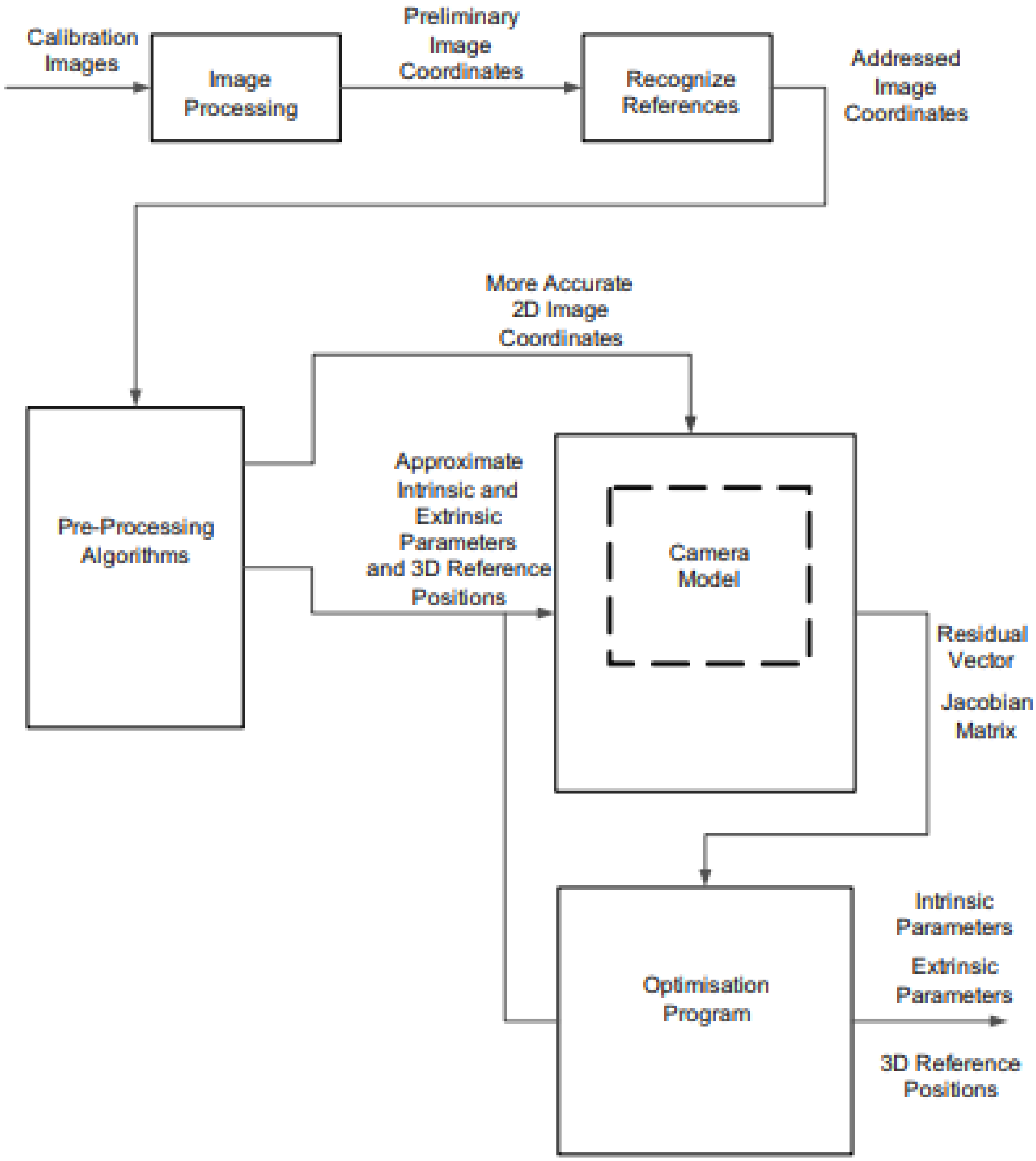

Figure 5: Camera Calibration [19]

7. OBJECT TRACKING

Anything of interest in a scene is known as object. Objects are recognized on the basis of their appearance and shapes. Points, lines, contours and skeletons give an idea of shape of an object. Various methods like object appearance probability densities, template matching, multi-view and active appearance models exist in the computer vision literature for representing features of an object. Object representations and tracking have a close relationship. For tracking very small objects like seeds or birds in a scene, generally point representation is preferred (Veenman et. al 2001). Square, rectangle, eclipse (Comaniciou et. al 2003) or eigen values (Black and Jepson 1998) are also used for object representation. A set of classifiers like SVMs, Bayesian etc. are trained to make the algorithms learn about different object views. [20]

Estimation of trajectory of an object in a scene is termed as object tracking. High power computer systems and inexpensive cameras have enabled scientists to track various object in a scene. In automated object tracking we assume smooth object path. Moreover we also assume that objects is moving with constant velocity and acceleration. Basic steps in video analysis are: detection of objects of interest, frame to frame path tracking and recognition of behavior of objects in the frame. [20]

7.1.1 Features

The features required for object detection and tracking are:

- **<u>Color:</u>**

Surface reflectance and spectral power distribution determine apparent color of an object in RGB, L*a*b, L*u*v or HSV color space. [20]

- **<u>Edges:</u>**

Changes in image intensities are indicated by edges. Boundary of an object is isolated from other objects using edges. [20]

- **<u>Optical Flow:</u>**

Translation of pixels in a region are indicated by displacement vector fields. In optical flow, we assume constant brightness. [20]

- **<u>Texture:</u>**

Smoothness and regularities are characterized by intensity variations in surfaces. Intensity variations are measured by texture. Illumination changes do not significantly affect texture features. [20]

7.1.2 Multiple Camera Tracking

To attain in depth object tracking information, increase field of view and reduce occlusions, the strategy of multiple camera tracking is employed. Interconnectivity between different camera views may be defined manually or computed automatically. Models of multi-camera tracking introduced by Mittal and Davis (2003) and Tekalp and Dockstader (2001) produce superior results as compared to the results, obtained by monocular camera tracking as they are subjected to occlusions. [20]

While object tracking, we have two assumptions:

i. we are using stationary camera and
ii. object track is available.

If the object being tracked follows the established path, we have high probability of finding superior results. [20]

If the object follows any arbitrary path, appearance and shape of object is recognized, the only approach applied is tracking by recognition. [20]

A taxonomy of object tracking methods is described below:

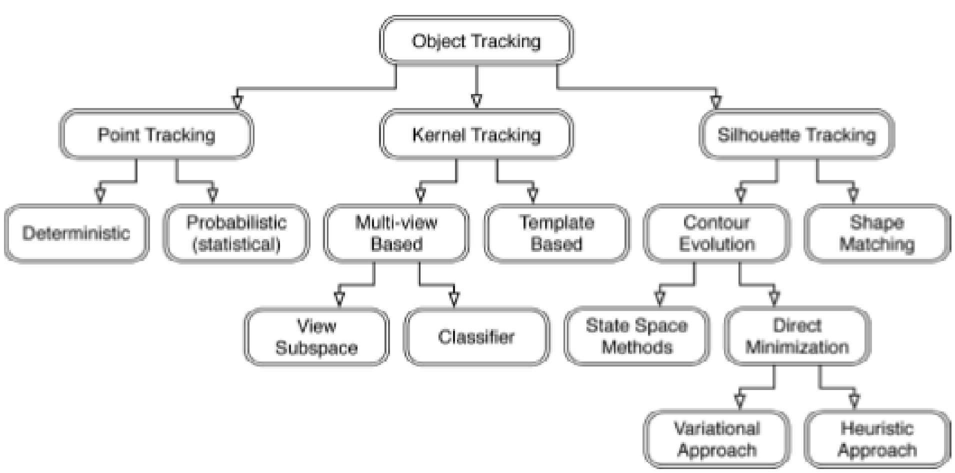

Figure 6: Object Tracking [20]

Table below shows various object tracking categories:

Table 1: Object Tracking Categories [20]

Categories	Representative Work
Point Tracking	
Deterministic Work	MGE Tracker (Salari and Sethi 1990)
	GOA Tracker (Veenman et al. 2001)
Statistical Work	Kalman Filter (Brodia and Chellappa 1986)
	JPDAF (Bar-Shalom and Foreman 1988)
	PHMT (Streit and Luginbuhl 1994)
Kernel Tracking	
Template and Density Based Models	Mean Shift (Comaniciu et al. 2003)
	KLT (Shi and Tomasi 1994)
	Layering(Tao et al. 2002)
Multi-View Appearance Models	Eigen Tracking (Black and Jepson 1998)
	SVM Tracking (Avidan 2001)
Silhouette Tracking	
Contour Evolution	State Space Models (Israd and Blake 1998)
	Variational Methods (Bertalmio et al. 2000)
	Heuristic Methods (Ranford 1994)
Shape Matching	Housdar (Huttenlocher et al. 1993)
	Hough Transform (Sato and Agarwal 2004)
	Histogram (Kang et al. 2004)

8. EDGE and LINE DETECTION

With the progress in the fields of Artificial Intelligence, Machine Vision and Deep Learning, a rapid development has been observed for network analysis, setting and manipulating systems for law enforcement agencies.

Every object is recognized by its particular shape when viewed by naked eye. For recognition of different objects in a scene, its shape has to be distinguished. The root of shape recognition lies in the location of edges and lines in an object. Shape is recognized by skeletonizing the object. Skeleton corresponds to set of points having a minimum of two points equidistant to the boundary of an object. [21]

- The first step in shape recognition is finding edges and lines in an image using some algorithm.
- The output image is imperfect with gaps and noise and often gives wrong edges.
- The inaccurate image is corrected by some image processing technique.

Analyzing an image indicates us about the lines and edges found in it. Shape recognition must classify all the lines found in the image. Image segmentation includes line and edge detection. **Region based segmentation** includes thresholding, splitting and merging. Discontinuity in a line indicates presence of edge. It is called **edge based segmentation**. Edges and lines are found by computing first and second order derivatives. [21]

Seatbelt is basically an object in the image, consisting of two parallel lines inclined at an angle between 50°- 70° as per standards. To get the detection done, we need to find seatbelt edges consisting of two parallel lines in the image, at least 80 to 100 pixels long. [6]

To go deep into the details, we need to have a good insight on ***edge*** and ***line detection***. This section presents an overview of common edge and line detection techniques prevailing so far.

8.1 Edge Detection

Edge is area in a digital image having pixels with intensity values different from neighboring pixels. Edge may occur due to variations in light intensity, color or texture. Analysis of images helps in filtering out irrelevant details for edge selection. [21]

Edge detection is a sequence of events required to identify the points in an image for which intensity levels change significantly. To extract information related to image i.e. object location, enhancement, image sharpening etc, this sequence of actions is required to extract the information regarding image. [21]

The basic step in edge detection requires a colored image. The image requires refining as much as possible without losing the important details. The image is then differentiated for enhancement in edge quality. Thresholding is done to ignore noisy pixels and those with useful details are confined. Sometimes sub-pixel resolution is required for edge position detection and finding the pixel spacing. After a series of fore mentioned steps we get the required edges in the image. [21]

Given ahead is a block diagram of edge detection algorithm.

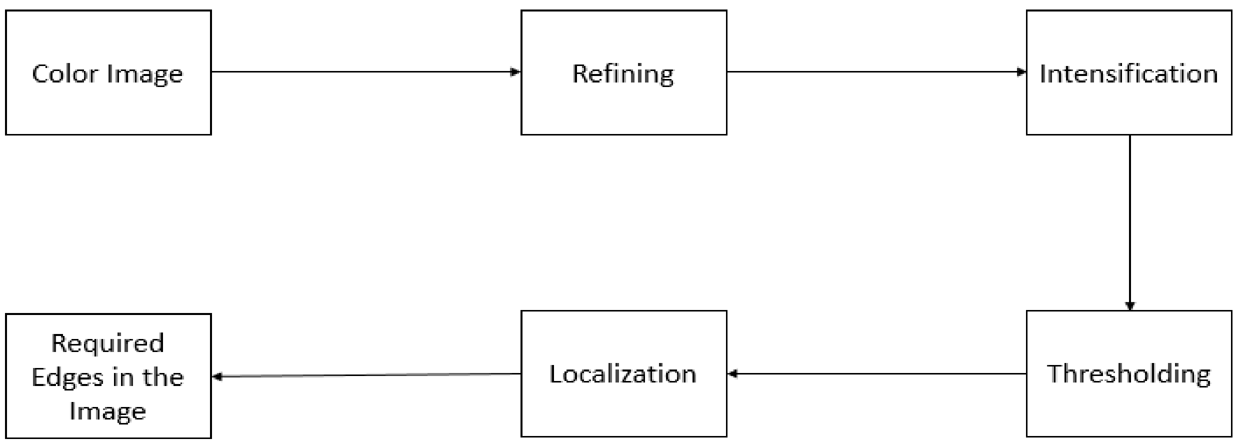

Figure 7: Edge Detection [21]

8.2 Approaches

We use Laplacian and Gradient Based methods in edge detection.

- **Gradient Based Edge Detection:**

  i. First order derivative of image is taken.
  ii. Edge strength is measured by gradient magnitude.
  iii. Gradient magnitude provides local maxima which helps in computing the orientation of local edge which lies in gradient direction. [21]

- **Laplacian Based Edge Detection:**

  i. Second order derivative of image is taken.
  ii. Edges are found from zero crossing of a non-linear differential equation.
  iii. Gaussian smoothing is applied in refining stage. [21]

The second order derivative's zero crossing expression avoids unimportant edges which correlate with first order derivative selected image edges of those pixel having value above some threshold. The Laplacian edge detector uses one mask for computing second order derivative mask. The LoG function and the "σ" standard Gaussian deviation is given as:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1-\frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{8-1}$$

There are three basic edge detectors:

- Sobel Edge Detector
- Canny Edge Detector
- Zero crossing Edge Detector

The properties of these detectors are given below:

Table 2: Sobel Vs Canny Vs Zero Crossing Detector [21]

Edge Detector	Basic Principles
Sobel	Finds edges using Sobel approximation to the derivatives.
Canny	Finds edges by looking for local maxima of f(x, y). The gradient is calculated using Derivative of Gaussian filter. The method uses two thresholds to detect strong and weak edges and includes weak edges only if they are connected to strong edges.
Zero Crossing	Finds edges by looking for zero crossing after filtering f(x, y) with a user defined function.

8.3 Line Detection

Knowing the location of edges, we can detect lines in an image. We know that infinite number of lines can pass through a point in time "t". Line detection algorithm takes "n" number of edge points thereby finding all of the lines on which edge points lie. [21]

8.3.1 Hough Transform

Hough transform is the most common method used to detect lines and outputs the parametric description of lines in the image, given by

$$\rho = r\cos(\theta) + c\sin(\theta) \qquad (8\text{-}2)$$

ρ is the equation of a line in image space based on column/row, the perpendicular distance from origin to line. θ is the angle of the projection from origin to line in degrees. Thus line in an image corresponds to a point in the Hough space. Hough space has two dimensions θ and ρ for lines. Each line has a unique θ and ρ. [21]

Hough transform is implemented by choosing certain values of ρ and θ. For each pixel in an image, we compute r cos (θ) + c sin (θ) and place the result in the appropriate position in array for ρ and θ. ρ and θ with the greatest values in the array correspond to strongest lines in an image. [21]

The main idea behind line detection in Hough Transform is:

- For every line, there exists a unique line which passes through the origin and is perpendicular to it.
- This line has a specific angle and distance from the “x-axis” of image.
- This angle and distance defines a point in Hough space.
- Infinite lines can pass through this point, having specific angle and distance.
- This set of lines corresponds to a sine function in Hough space.
- Two such points on that line in the image correspond to two sinusoids crossing at a point in Hough space.
- This point in Hough space corresponds to a line in image space.
- All sinusoids on that line will pass through that point.

Hough space is quantized using 2D counter array called accumulator array. Line detection in Hough transform has three stages:

- Perform accumulation on the **accumulator array**.
- Finding **peak** values in that accumulator array.
- Verifying that the peaks indicated are corresponding to legitimate lines not the noise.

To segment an image into its components, image segmentation is performed using **global thresholding**. The sequence of events in thresholding are listed below:

i. Initially we select an estimate for a threshold “T”.
ii. Segment the image with values of “T”.
iii. It gives two sets of image pixels, one with pixel intensity greater than “T” and other with intensity less than “T”.
iv. Average intensity of pixels of both groups is computed.
v. A new averaged threshold “T” is calculated.
vi. Steps 2 and 5 are repeated till we get “T” less than initial estimate “T”. [21]

9. COLOR CLASSIFICATION and SEGMENTATION

Color is a crucial factor in computer vision and image processing applications. It plays a vital role in object recognition, segmentation and detection. Various color models are available in literature that provide relative information about images and videos. These models can be transformed into one another upon requirement. [22]

Given below is a brief survey of various color models.

9.1 Color spaces

Color space is a coordinate system within that system in which every single color is denoted by a single point. It facilitates the specification of colors. This representation in 3D provides a color manipulation information and is a natural way of representing a spatial relationship among various colors. These models are capable of transformation from one space to other. [22]

9.1.1 RGB Model

RGB model is the most commonly used model. It is widely used system in digital images and is compatible with a wide variety of color cameras. [22]

- In RGB model, every color is represented in terms of primary constituents red, green, and blue outputs (R, G, B) from a color imaging system. The model is based upon a Cartesian Co-ordinate System.
- RGB values are at three corners of a cube; yellow, magenta and cyan are at other corners, representing color sub-space of interest.
- Black lies on the origin and white is present at the corner farthest from the origin. Along the line joining these points, gray scale values extend from black to white.
- The RGB values are sensitive to intensity and color changes. [22]

The figure below shows an RGB color cube.

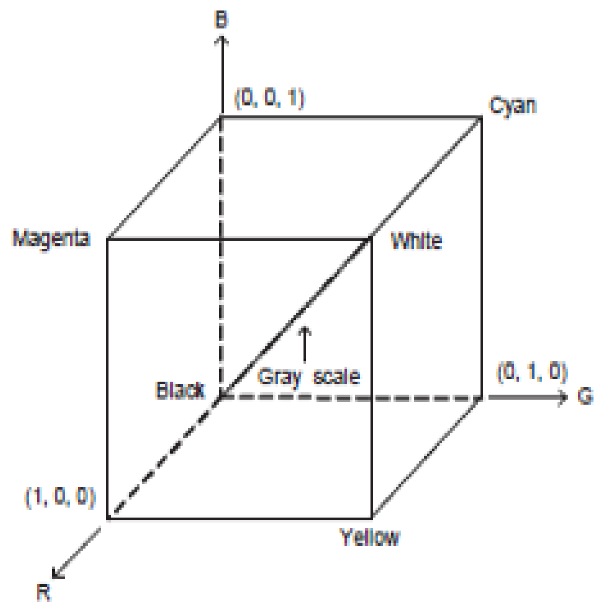

Figure 8: RGB Color Cube [22]

9.1.2 HSV Model

A color object is represented by its hue, saturation, and intensity. HSV stands for hue, saturation and value. [22]

- **Hue** describes a pure color. Hue is also defined as the dominant wavelength of a dominant color of light viewed by an observer e.g. the blue car or green colored leaf.
- **Saturation** gives the measure of dilution by white light.
- **Value** or gray levels is also a key factor to describe a color and is easily interpretable. [22]

Figure below shows the components of an HIS image.

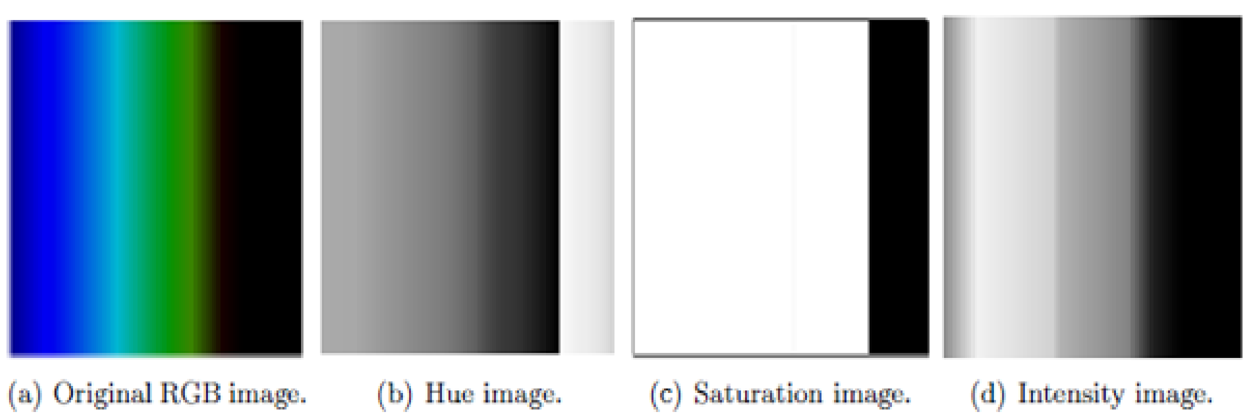

Figure 9: HIS Image [22]

9.1.3 XYZ Color Model

To express a complete range of colors seen by human eye, researchers defined a color space named XYZ Color Model. [22]

XYZ color space needed to fulfill following criteria:

- Three positive and independent variables are required sufficiently to specify a color.
- Just tri-stimulus source values were relevant to express any color, the spectral composition was not needed.
- Even for one source value changed gradually, tri-stimulus values would vary gradually as a result. [22]

All the colors of visible spectrum are represented in positive X, Y and Z value combination.

The XYZ color space expresses every color visible to human eye, which means that it is capable of expressing all the colors being captured by a camera, hence every color that anyone might ever want to reproduce. Thus DCI (Digital Cinema Initiative) adopted XYZ Model as its standard color space. While possible to make a camera close to recording colors in the XYZ space, "a perfect camera", building a monitor working in the XYZ color space is not possible as X, Y, Z are not real colors of light. [22]

It particularly is not possible to produce a range of blue-green colors – that can perfectly be captured in a camera. The reason of color loss is the response cones present in eye to differentiate different wavelengths of light (red, green, blue). The x (λ), y (λ) and z (λ) curves outline the responses of the red, green and blue cones respectively. These curves show how each type of cone responds to the wavelengths of the visible spectrum. [22]

The figure below shows eye response diagram regarding selection of red, green and blue primaries.

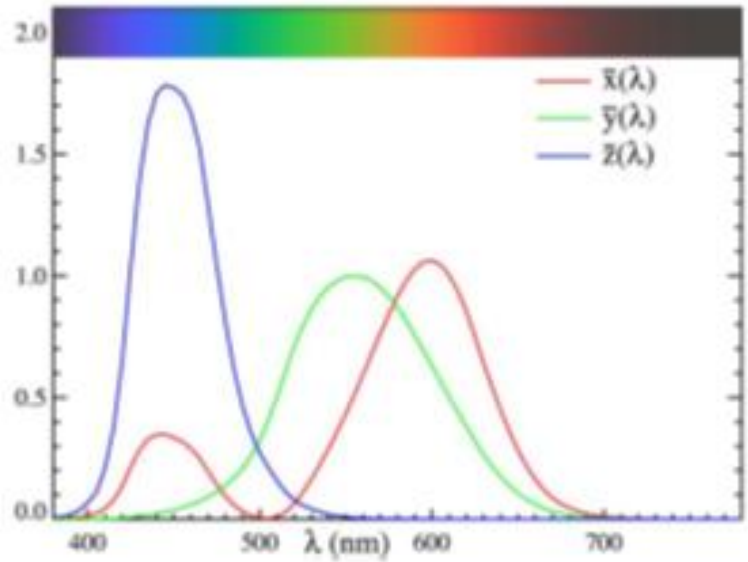

Figure 10: XYZ Color Model [22]

A particular component of wavelength induces a signal in the red, green and blue cones, equals to intensity of the light multiplied with wavelength of light at that instant to produce cone's response to light of that wavelength. [22]

The response can be expressed mathematically as:

$$X = \int I(\lambda).\bar{x}(\lambda)d(\lambda) = \ \textit{Signal from Red cones}$$

$$Y = \int I(\lambda).\bar{y}(\lambda)d(\lambda) = \ \textit{Signal from Green cones}$$

$$Z = \int I(\lambda).\bar{z}(\lambda)d(\lambda) = \ \textit{Signal from Blue cones} \qquad (9\text{-}1)$$

Any tri-stimulus value of one color space can thus be converted corresponding tri-stimulus value of another color space by applying a linear transformation. XYZ and RGB value inter-conversion by linear transformation is represented mathematically as:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (9\text{-}2)$$

Similar transformation can be applied in conversion from RGB to the XYZ color space. [22]

10. CAMERA TYPES and INTERFACES for MACHINE VISION

Machine vision systems contain several components, like lenses, camera, acquisition and measurement, data transfer and image processing software. Selection of optimum components for a robust system should be based on the ability to achieve the required measurement in the presence of external environmental influences. [22]

Machine vision applications in the manufacturing, assembly, biotechnology, electronics, and semi-conductor industries where analytical inspection is required, use of best camera system for is significant to achieve best image quality. From digital and analog cameras to progressive scan to GigE and FireWire interfaces, parameters like camera types, digital interfaces, power, and software provides a great opportunity to upgrade one's self to imaging expert. [22]

10.1.1 Analog Vs Digital Cameras

Cameras are divided in two types on general level: analog and digital.

In real time**, analog cameras** transmit a continuous, variable electronic signal. The amplitude and frequency of attained signal is interpreted later on by analog output device as video information. Analog video quality and the way of interpretation, both affect the resulting video images.

The pros and cons of this data transmission method exist.

- Analog cameras are less complicated and less costly than digital cameras, making them cost-efficient and simple for usual video applications.
- Analog cameras however have limitation for resolution and frame rate, both.
- Analog cameras are subjected to electronic noise, due to overlooked factors like cable length and connector type. [22]

Digital cameras transmit a sequence of ones and zeroes (binary data) in the form of an electronic signal.

Although voltage level corresponding to the light intensity of a pixel is continuous, the AD conversion process assigns grayscale values between 0 (black) and 2^{N-1}, here N represents number of bits for encoding.

Output device converts binary data into video information later on. Important key differences uniquely present in digital cameras types are:

1. The digital signal is the same when it leaves the camera and reaches an output device.

2. Digital video signal interpretation is done in one way.

- The differences remove errors in signal transmission and interpretation by output device due to the display. As compared to analog cameras, digital cameras offer high resolution and frame rates, more features and less noise. [22]
- The advantages bear cost - digital cameras are more expensive in general. Moreover, feature-packed cameras involve complex setups, even for basic capabilities. These cameras are mostly limited to shorter cable lengths. [22]

10.1.2 Digital Camera Interfaces

Signal degradations like distortion, transmission noise etc. don't affect the information transmission through a digital camera. Due to digital signal, information loss is less in the transmission process. The commonly available interfaces these days are FireWire , capture boards, GigE, Camera Link, and USB. The choice of suitable interface depends upon the end user requirement.

Due to two way communication, asynchronous transmission allows data transfer receipts, ensuring signal integrity and in time delivery. For isochronous transmission, scheduled packet transfer occur at pre-defined time like every 125µs, ensures timing but at high transfer rates, allows the possibility of dropping packets. [22]

10.1.2.1 Capture Boards

Use of computers is involved in image processing. Capture boards contain an ADC for digitizing the signal, used later for image processing. Purpose of capture boards is to allow users to output camera signals into computer for analysis purpose. [22]

Capturing software is provided with capture boards to allow users for saving, opening and viewing images. Capture boards also refer to PCI cards, necessary for acquiring and interpreting data from digital camera interfaces. [22]

10.1.2.2 FireWire (IEEE 1394)

FireWire is serial, ***isochronous camera interface*** pertaining to availability of ports for FireWire on computers. Firewire.a provides ***slower transfer rates***, yet both FireWire.a and b allow connection for multiple cameras, providing power through FireWire cables. [22]

10.1.2.3 Camera Link

It is a ***fast-speed serial transfer interface*** typically developed for computer vision applications. Capture card is needed for using camera link and separate power is required to be supplied to camera. Provision of ***separate serial communication channels (asynchronous)*** to retain complete bandwidth is done. Special cabling is needed in addition to LVDP (low-voltage differential pair) signal lines. [22]

Single cable base configuration dedicated for video, allows 255 Mbps data transfer rate. Dual (full) cable base configuration provides separate sending/receiving lines, providing higher data transfer rates up to 680 Mbps in high speed requirements. [22]

Camera Link incorporates optical fiber cable support up to 300m length. [22]

10.1.2.4 GigE Vision

Gigabit ethernet protocol uses Cat-5&6 cables for providing higher speed interface for camera. Considering overall bandwidth, hardware like hubs, switches and repeaters are used for multiple camera interfaces.

(IEEE 802.3 ad) Link Aggregation uses multiple parallel ethernet ports for increasing data rates and multicasting for distributing load on processor. The PTP (Precision Time Protocol) supported by some cameras, is used for synchronizing multiple camera clocks connected on one network thus allowing a definite predictable delay relationship. [22]

10.1.2.5 USB (Universal Serial Bus)

USB 2.0 is a commonly used interface. It is convenient but provides medium speed. The data transfer rate is fixed at 480Mbps. Maximum attained speed usually depends on number of USB peripherals.

Sometimes (usually with laptops), it is needed to apply separate power to the camera. [22]

USB 3.0 has a feature of plug-and-play and allows for comparatively higher data rates.

10.1.3 Camera Software

There are two choices for an imaging software: third-party software or SDKs (camera-specific software development kits).

SDKs include application programming interfaces with code libraries for development of user defined programs, as well as simple image viewing and acquisition programs that do not require any coding and offer simple functionality.

Third party software includes MATLAB, NI LabVIEW, OpenCV etc. Third-party software can support multiple interfaces and multiple cameras, but the user needs to ensure the functionality. [22]

By understanding pros and cons depending upon application, user can pick the combination of choice for an application with respect to interfaces, camera types, power requirements, and software. [22]

The table below illustrates a comparison of some digital camera interfaces.

Table 3: Digital Camera Interfaces [22]

Digital Signal Options	Fire 1394.b	Camera Link	USB 2.0	USB 3.0	GigE
Data Transfer Rate	800 Mb/s	3.6 Gb/s (full configuration)	480 Mb/s	5 Gb/s	1000 Mb/s
Maximum Cable Length	100m	10m	5m	3m (recommended)	100m
Connector	9 pin- 9pin	26 pin	USB	USB	RJ45/Cat5e or 6
Capture Board	Optional	Required	Optional	Optional	Not Required
Power	Optional	Required	Optional	Optional	Required

11. CAMERA SENSORS

Both “imaging electronics and optics” have an important role in imaging system’s performance. Integrating required components i.e. camera, software and capture board form an optimal imaging system. [22]

Sensor is the heart of any camera. Solid state sensors contain numerous discrete photo-detector sites called pixels. Due to varying design of interfacing electronics, two cameras having similar sensor can have very different performance. Today, CMOS and CCD constitute nearly all of the machine vision imagers. [22]

11.1.1 Charged-Coupled Devices (CCD)

The CCD sensor actually is a silicon chip containing photosensitive sites. Clock pulses tend to create potential a difference for moving charge packets on the chip, before they are converted into voltage by a capacitor. ADC converts output of analog CCD into a digital signal in digital cameras. [22]

Voltage from every photo detector site, in analog camera, is read in a specific sequence, with added synchronizing pulses for image reconstruction, in signal chain. [22]

Because of same voltage conversion, CCD sensor is highly sensitive and pixel-to-pixel consistent. CCD is very much uniform across photo sensitive sites. [22]

11.1.2 Complementary Metal Oxide Semiconductor (CMOS)

The charge from photo receptor pixel is converted into voltage in a CMOS sensor. The corresponding signal is multiplexed to on chip DACs, by rows and columns. CMOS itself is a digital device and performs the functions of pixel activation, charge conversion, amplification and multiplexing. This leads to the less sensitive, high speed CMOS sensors and possess high noise of fixed-pattern due to inconsistent fabrication in the multiple charge to voltage conversion circuits. [22]

An additional advantage of a CMOS sensor is its high efficiency and less power consumption compared to CCD sensor, due to less charge flow. Ability of CMOS sensor to handle high intensity light without blooming makes it suitable for range of special and highly dynamic cameras. Moreover, a digital CMOS camera is smaller than its CCD counterpart because CCD digital camera needs off-chip ADC circuitry in addition. [22]

Multilayered MOS fabrication of CMOS sensor doesn’t allow use of micro on-chip lenses, decreasing effective collection efficiency of the sensor compared to an equivalent CCD sensor. Low efficiency thus combined with pixel-to-pixel inconsistency leads to a lower SNR and overall image quality than CCD sensors. [22]

The below figure shows CCD and CMOS.

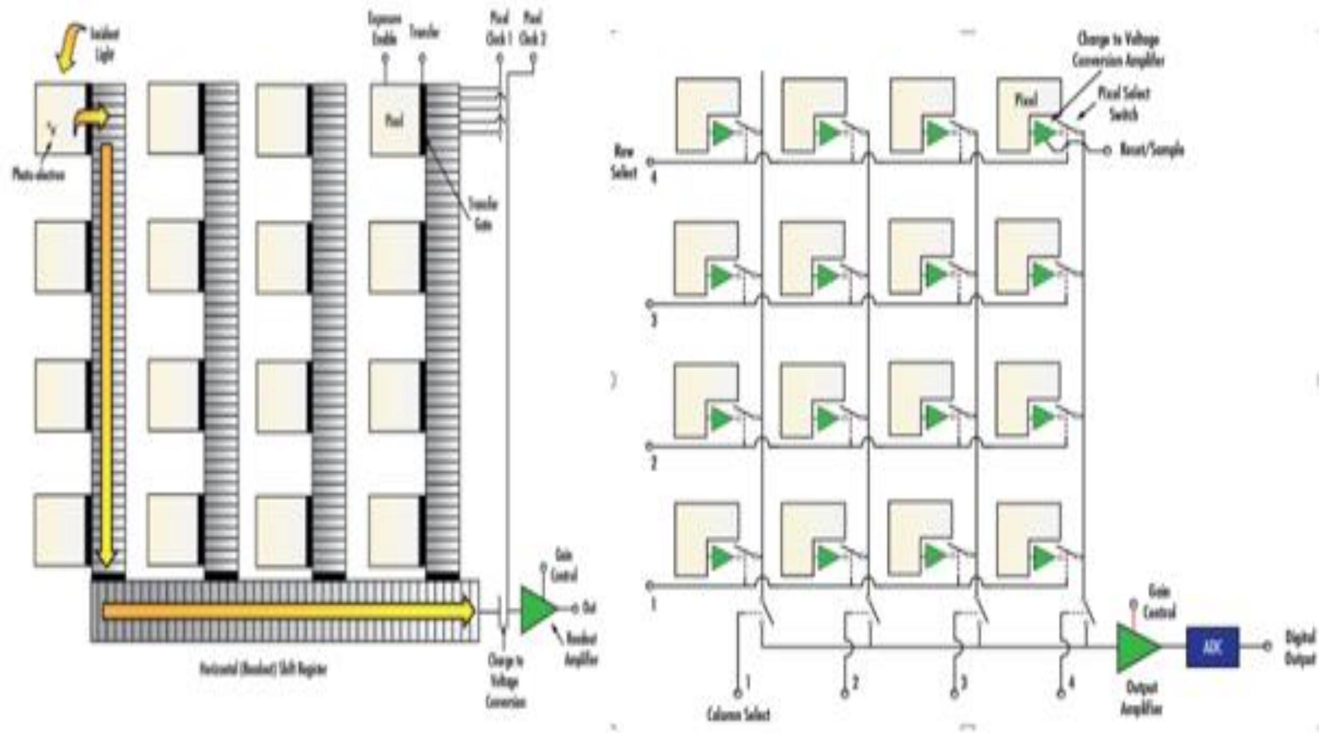

Figure 11: CCD and CMOS [22]

11.1.3 Sensor Features

11.1.3.1 Pixels

A matrix of pixels collect light falling on a camera sensor. Information of these photo receptor sites is collected, organized, and transferred to a monitor for display. The pixels may be photodiodes or photo capacitors generate a charge, proportional to the amount of light incident on them. Quantum efficiency specifies trait of that pixel, converting an incident photon to charge. Typical range of quantum efficiency for solid-state imagers is between 30 - 60%. Wavelength of light specifies its quantum efficiency and is non-uniform light intensity. Spectral response curves specify quantum efficiency in terms of function of wavelength. [22]

Common pixel sizes are between 3 - 10μm. A general measure of sensor's resolution is number of pixels, saturated per millimeter. The size of pixel is considered important in imaging optics. Large pixels usually have high SNRs and charge saturation capacities. Small pixels usually have high magnification and resolution. [22]

CCD analog cameras have rectangular pixels. Asymmetrical pixels provide high horizontal resolution as compared to vertical. CCD analog camera have same vertical resolution. Camera industry specifies horizontal resolution as a standard. [22]

11.1.3.2 **Sensor Size**

The size of sensor's active area is crucial to determine its field of view. Large sensors have larger FOV. Several standard sensor sizes exist: ¼"", ½", 1" and 1.2" etc. Most of the sensors maintain a (Horizontal: Vertical) 4:3 dimensional ratio. [22]

One of issues that can arise is the ability of a lens to support specific sensor sizes. The image may fade away at the edges if sensor is larger for lens design because of extinction of rays passing through outer edges of the lens. This is called tunnel effect. Small sized sensors do not produce this issue. [22]

11.1.3.3 **Frame Rate and Shutter Speed**

Frame rate is defined as number of frames per second. In high-speed applications, faster frame rate is required for acquiring maximum images of an object as it moves through the field of view. [22]

The exposure time of the sensor represents its ***shutter speed***. The amount of incident light is controlled by exposure time. ***Camera blooming*** is **controlled** by ***lowering illumination or raising the shutter speed***. [22]

For approximation, taking ***inverse of the frame rate gives shutter speed***. Due to resetting pixels and reading out of pixels, there exists a minimum time of the order of microseconds between exposures. This time can be taken out from the camera datasheet. [22]

11.1.3.4 Electronic Shutter

Previously, CMOS cameras used rolling shutters and CCD cameras used electronic shutters. In an ***electronic shutter***, all pixels are sampled simultaneously and for all pixels, photon acquisition starts and stop time is same.

A ***rolling shutter*** reads and samples out sequentially; there is a different sample time for each line. Rolling shutter distorts the images of moving objects. This can be decreased by a triggered strobe, placed at the point where the overlapping of integration period of lines take place. Implementing electronic shutter for CMOS needs a more complex architecture than rolling shutter model, with an added storage capacitor and transistor for pipelining. High-speed motion applications can use both CMOS as well as CCD cameras. [22]

In comparison to rolling and electronic shutters, asynchronous shutter relates to the triggered exposure of pixels i.e. the camera acquires the image after receiving external trigger signal. This effect is opposite in normal constant frame rate, which receives an internal triggering for shutter. [22]

11.1.4 Spectral Properties

11.1.4.1 Monochrome Cameras

CMOS and CCD sensors are sensitive to 400 - 1000nm wavelength. Sensor's spectral response curve of sensor indicates the sensitivity. IR cut-off filters are provided in most high quality cameras for imaging. [22]

Generally, CCD sensors are less sensitive for wavelengths of IR region than CMOS sensors. This is because of their decreased active area depth. Photon's penetration depth depends on its specific frequency. Deeper depths for active area thickness decreases quantum efficiency and produces less photoelectrons. [22]

Normalized Spectral Response of CCD is shown below.

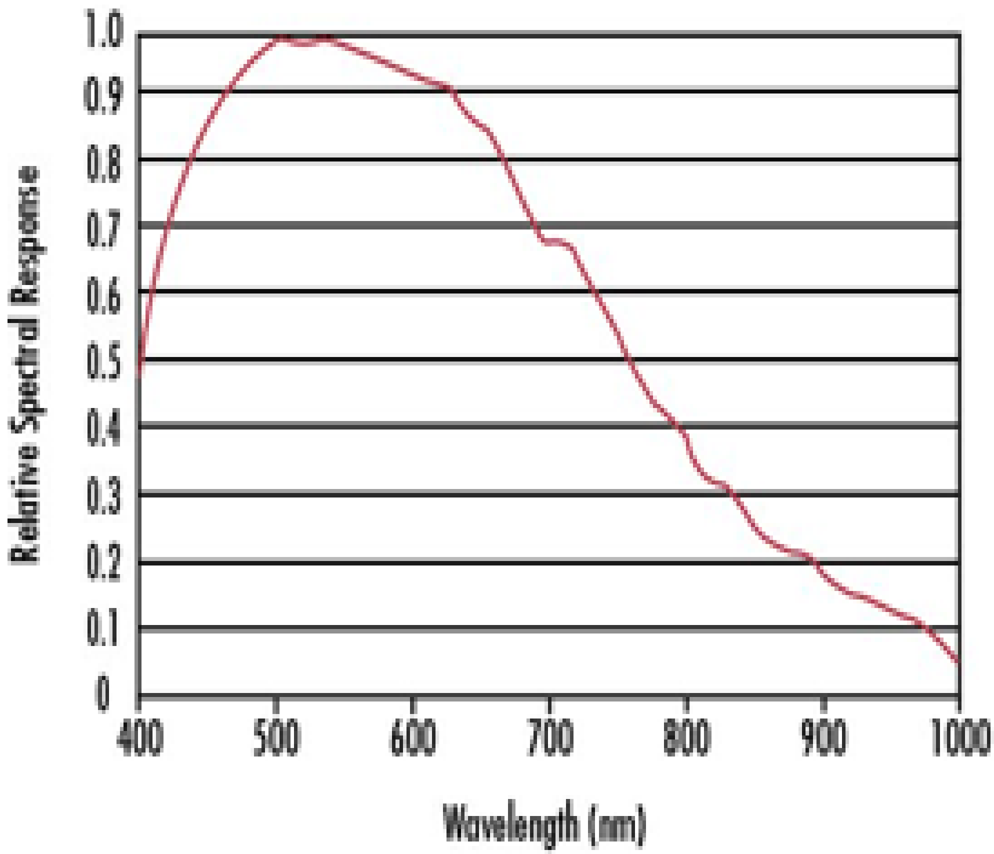

Figure 12: Normalized Spectral Response of CCD [22]

11.1.4.2 Color Cameras

Photo-electric effect forms the basis of solid-state sensor, therefore it can't distinguish colors. CCD cameras have two types:

- Single chip and
- Three-chip.

Single chip CCD camera offers less costly imaging solution, uses filter e.g. Bayer filter for segregating incident white light into seven constituent colors. Each color corresponds to a specific pixel set. Since, color recognition requires more pixels, single chip cameras subsequently have low resolution than monochrome cameras. Color interpolation algorithm specified by manufacturer does color recognition. [22]

Three-chip CCD camera solves color ***resolution problem*** by directing section of the incident spectrum to a specific chip using a prism. Much accurate color reproduction is possible, because every point in color space of object has specific intensity of RGB values and does not use a color recognition algorithm. Three-chip cameras are costly but offer very high picture resolutions. [22]

12. DATA DESCRIPTION

One of the main challenges of the study was the availability of appropriate dataset for driver seatbelt. Many of the datasets like Microsoft COCO, YOLO etc. prove helpful in general object detection. No specific dataset is available for seatbelt detection.

To train our system as close to the real time scenario as possible, we collected our dataset on three different locations on Attock- Rawalpindi Motorway M-1 and on GT- Road Rawalpindi on different overhead bridges. Multiple clips of different timing with 24 fps and 60 fps were taken. The cameras had an overhead view in the same fashion as Motorway Police Department does with its cameras for finding vehicular speed. The locations and characteristics for video imaging were selected considering driving conditions in Pakistan. Grid References (GRs) are taken with respect to Indian Zone 1 using GP Seismic Software.

The table below, summarizes locations, characteristics and GRs of places used for data collection:

Table 4: Location Information of Collected Data

Sr. No.	**Type**	**Description**	**Grid Reference (GR)**
1	**Motorway M-1**	Overhead bridge 5.2 km from Islamabad Interchange	912508
2	**Motorway M-1**	Overhead bridge 1km from Wah-Taxilla Interchange	712702
3	**Highway**	GT Road overhead bridge opposite to CEME Rawalpindi	022496
4	**Highway**	Rawalpindi-Peshawar GT Road (Haji Shah Overhead Bridge)	426765

13. METHODOLOGY and RESULTS

The prevalent speed measurement method in Pakistan is quite orthodox. The traffic police department computes the velocity of a vehicle using LIDAR gun, operated manually. No method of seatbelt detection exists till date. What we are trying to do is to design an automated device for detection of driver's seatbelt and vehicle speed all through a video frame using machine vision.

The device hardware consists of:

- Camera and
- Computer system

Camera-Computer interface utilizing OpenCv Python is used to determine driver's seatbelt and vehicle speed using Haar Cascade Classifier. The video-feed can be input from any of the saved videos in computer and camera can also be connected to the computer for live demonstration.

13.1 Methodology

The task is divided into following four sub-tasks:

13.1.1 Designing a Haar Cascade Classifier

- Haar Cascade is used primarily to perform object recognition in the scenes, image classification and clustering them based on similarity.
- Our task requires classification of vehicle, driver and seatbelt in a video frame.
- Based on the video-feed of monocular camera, we classify objects of our requirement in the scene.

13.1.2 Training Haar Cascade on the Dataset

Training the Haar Cascade on our own dataset enables the system to classify the required objects in the scene. Some important parameters of Haar cascade training include the following:

i. Vector file containing positive samples,
ii. Vector file containing negative samples,
iii. Number of positive and negative samples,
iv. Number of stages used for training,
v. Maximum false alarm rate of each stage,

vi. Minimum hit rate of each stage,
vii. Splits in each stage,
viii. Symmetric or Non-symmetric object,
ix. Mode of features to be used (BASIC, CORE, ALL)

- Two separate Cascade classifiers were made with different number of training images.
- "Cascade-Seatbelt (.xml file)" had 464 positive images and 1039 negative images and"Cascade-Seatbelt1 (.xml file)" had 226 positive images and 452 negative images for training our cascade.
- Training parameters of Haar cascade files are depicted below:

Table 5: Cascade Training Parameters

Parameters	**Cascade-Seatbelt**	**Cascade-Seatbelt1**
Training Image Size	32x32	32X32
Stage Type	BOOST	BOOST
Feature Type	HAAR	HAAR
Boost Type	GAB	GAB
No. of Stages	20	20
Positive Training Images	464	226

13.1.3 Speed Measurement

- The Cascade interprets vehicle speed between two points located at a fixed distance.
- The system computes speed by measuring the time required to cross that distance by the vehicle.

The steps involved in speed computation are:

a) vehicle detection and applying bounding boxes on each vehicle,

b) assign new tracker ID to each vehicle,

c) camera angle was not purely horizontal but is calculated to be 3.2° considering camera height 7.0104 m and distance of image plane to be 125 m,

d) calculate camera calibration coefficient from the known parameters i.e. perpendicular view (P), distance between object and camera (D) on the basis of camera height, view angle of camera ,angle of camera and image parameters.

$$P = 2\tan\left(\frac{Tc}{2}\right)\sqrt{H^2 + D^2}$$

$$K = \frac{P}{I_h} \qquad (13\text{-}1)$$

In our case, camera calibration coefficient is 8.5,

e) perform vehicle tracking in successive frames using "*dlib correlation tracker function*". "*Correlation Tracker function*" requires an array of Gray scale or RGB image. This function tracks vehicle in a bounding box in subsequent frames and updates itself constantly with new position of vehicle under track in consecutive video frames, for attaining distance,

f) attain eucleidian distance in meters from total tracked distance,

$$\text{d\_pixels} = \sqrt{new\ position^2\ - previous\ position^2}$$

$$\text{d\_meters} = \text{d\_pixels} / k \qquad (13\text{-}2)$$

g) convert speed to km/h by using formula:

speed = d_meters * fps * 3.6 (13-3)

Note: Factor of 3.6 here, means conversion of m/s to km/h, 1km= 1000m therefore 1m= 1/1000 km and 1h= 3600 s therefore 1s= 1/3600 h so 3600/1000= 3.6

13.1.4 Seatbelt Detection

- First step in seatbelt recognition is the detection of vehicle in a video frame and then locating seatbelt area in the upper body of driver.
- Steps of seatbelt detection are summarized below:

i. Access pre-trained ".xml" file from a known path,

ii. Access video for seatbelt detection from a known path,

iii. Reset original video for better detection,

iv. Covert RGB to Gray scale image,

v. Set magnification and nearest neighbors,

vi. Detect seatbelt in the upper body of driver's area from vehicle according to positive training images,

vii. Put bounding box around the detected seatbelt region,

viii. Save the detected images in a folder of known path.

Haar Cascade:

i. assigns weights to training dataset,

ii. selects a weak classifier and subsequently moves from weak to strong classifier,

iii. for all positive and negative samples, computes eigen values and weighted error rates,

iv. classifier with smallest weighted error is chosen as a weak classifier,

v. ultimately gets a strong classifier after repeated iterations.

The image below shows a sample of training images for Haar Cascade:

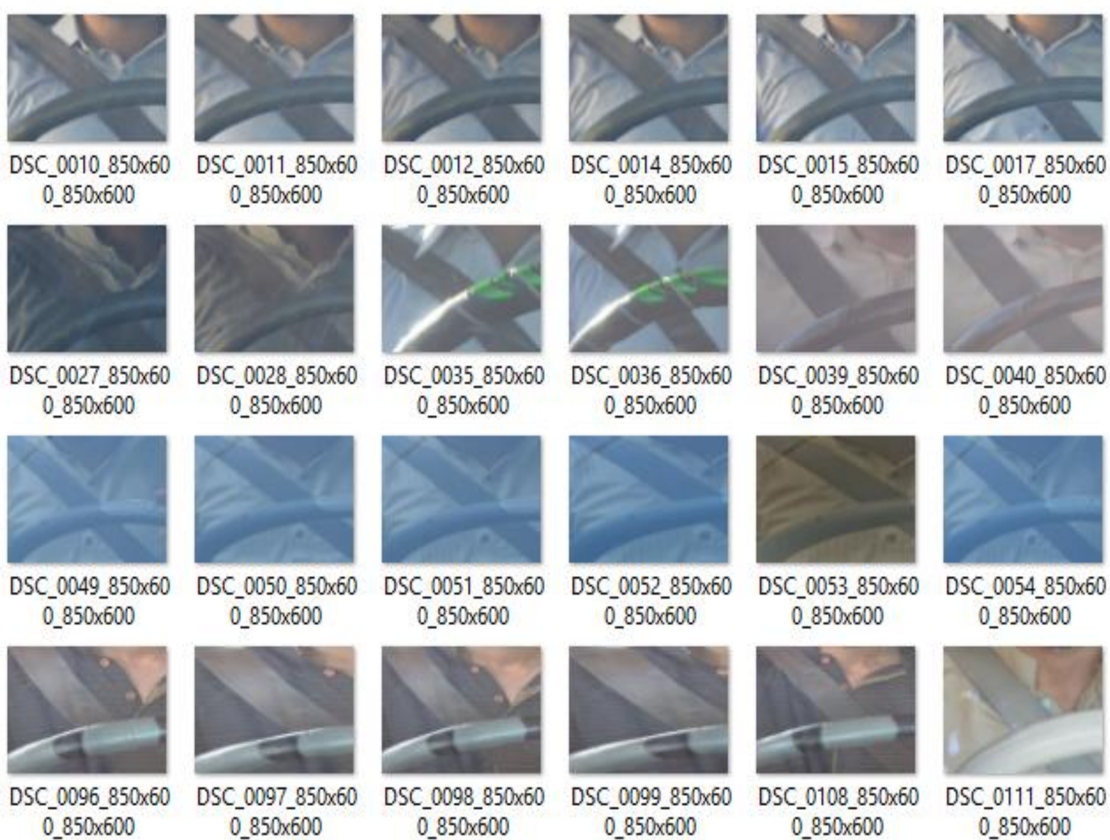

Figure 13: A Sample of Training Images

13.2 **Results**

The system is designed considering the driving parameters in Pakistan. The lane width of road varies from area to area but highways and motorways are of particularly larger width. We used a static monocular camera to record video frames to be used in detection. It was assumed that average lane width of motorway M-1 is 12 feet, tracking pathway is fairly flat and camera had zero distortion. All the experimentation process was done on" hp core i5 7th generation laptop".

13.2.1 Speed

Speed is detected by vehicle tracking along the path from point of entrance to point of exit from video frame. A vehicle is required to travel a distance between 100-200 meters for accurate results. Single camera per lane is not required. All the lanes of motorway or highway can be covered with a single camera but precise camera calibration is required.

Video was captured on Rawalpindi- Peshawar GT Road (Haji Shah Overhead Bridge) at 60 fps.

To find the variance between actual speed and detected speed, a vehicle with known speed was used and result is incorporated in the table.

A brief comparison of Actual Vs Detected speed of vehicle is presented below:

Table 6: Actual Vs Detected Speed

Sr. No.	**Actual Speed (km/h)**	**Measured Speed (km/h)**	**Variance (km/h)**
1	50	48	2
2	70	63	7
3	80	72	8
4	90	82	8
5	100	110	10

Graphical interpretation of data is depicted as under:

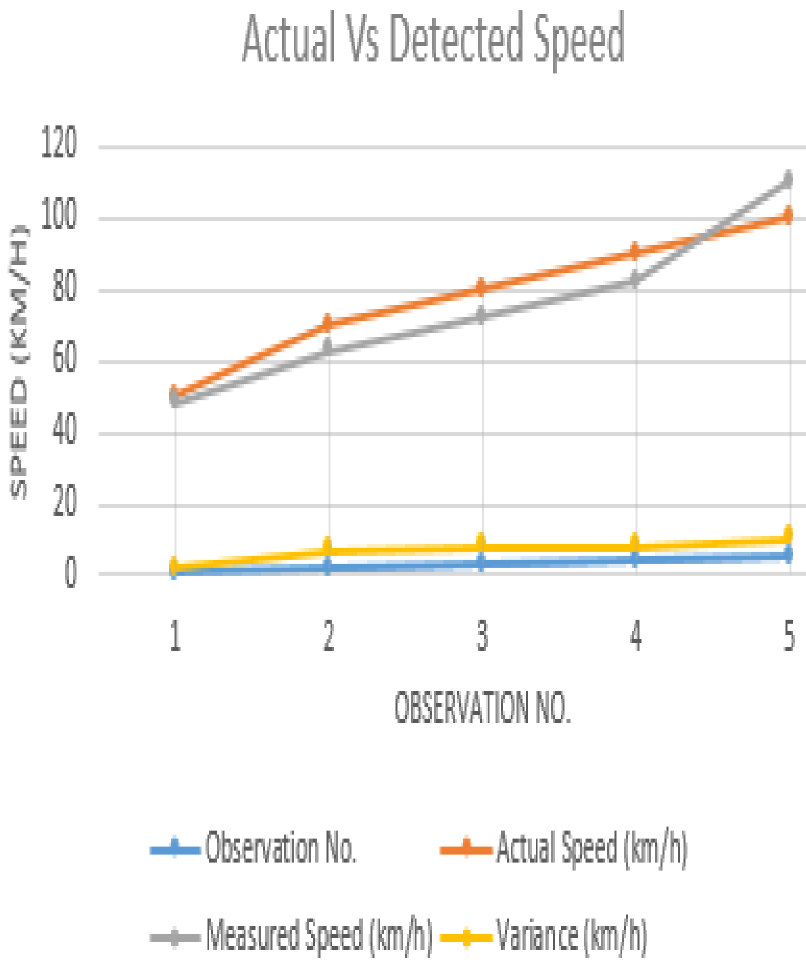

Figure 14: Graphical Interpretation

Comparing the results with "Detection of 3D Bounding Boxes of Vehicles Using Perspective Transformation for Accurate Speed Measurement", published in March 2020 by V. Kokur, M. Ftacnik, the maximum attained precision was 89.26% while our results are precise by 90%.

13.2.2 Seatbelt

The program looks for possible seatbelt area on upper body in the video frame and assigns a bounding box on that detected area. The videos were recorded on Attock-Rawalpindi Motorway M-1 on over-head bridges at 24 fps.

Horizontal view of moving vehicle with respect to camera works best for seatbelt detection.

Best results are obtained when we capture single lane with a camera, accuracy of results decreases when multiple lanes are captured with single camera.

Comparing the results with “Detection and Implementation of Driver’s Seatbelt on FPGA” published in Wu Tianshu et al. in 2019, results were precise by 81% while our results are precise by 87%.

A brief analysis of results is mentioned below:

- The model works well for very clear and semi-clear view of driver in correspondence with seatbelt detection. In the cars with dark gray or black windshields, it gets very difficult to detect whether driver is wearing seatbelt or not.
- In a few cases for heavy duty transport vehicles including trucks and dumpers, tires are also identified as seatbelts as an error detection.
- In some cases, side view mirrors of driver are also identified as seatbelt area as error detection.

- With respect to the number of vehicles detected, video “Motorway1” of 10 minutes duration taken at GR “712702”, 30 vehicle drivers were wearing seatbelt and algorithm detected 32 with two false positives for 140 percent magnification and 6 nearest neighbors. Another video “Motorway7” taken at GR “712702”, 28 vehicle drivers were wearing seatbelts and algorithm detected 32 with four false positives for 140 percent magnification and 6 nearest neighbors.
- Precision is calculated by formula:

$$\text{Precision} = TP / TP+FP \qquad (13\text{-}4)$$

TP represents True Positive and FP represents False Positive image.

Table 7: Results

Parameters	**Cascade Seatbelt**	**Cascade Seatbelt1**
Video Title	Motorway1	Motorway1
Magnification	140%	140%
Nearest Neighbors	6	6
True Positive (TP)	248	222
False Positive (FP)	30	39
Precision	89.2	85

Figure 15: Seatbelt Detection

Figure 16: Seatbelt Detection

14. CONCLUSIONS

The thesis introduced a device to yield vehicle speed and detection whether driver seatbelt is ON or not, using Haar Cascade Classification.

A brief introduction, statistics, objectives of research and brief description of proposed solution was presented in Chapter 1.

Chapter 2 introduced a brief literature review and related work regarding driver's seatbelt and speed estimation. Chapter 3 discussed four prevalent devices for manual speed estimation and the one used by Pakistan's Motorway and Highway Police is LIDAR gun.

Chapter 4 described Haar Cascade Classification and its insight. Chapter 5 explained Object detection and ways to detect objects in a scene. Chapter 6 elaborated Camera calibration.

Chapter 7 introduced Object tracking, Features for object tracking and Relevant Work regarding Object tracking categories. Chapter 8 explained Edge and Line detection, Various approaches like Laplacian and Gaussian based methods. Chapter 9, 10 and 11 explained Color space models, Camera Types, Interfaces for Machine Vision and Camera Sensors for machine vision appilcations.

Chapter 12 explained the Procedure and location information regarding collected data. Chapter 13 explained Methodology and results.

The system was trained on our own data set for seatbelt detection and pre-trained model was used for speed detection. Vehicle tracking is required in both speed and seatbelt detection and system optimally goes from weak to strong classifier.

The results obtained are quite promising indeed. The main advantage of the device is low cost, good accuracy and acceptability.

The system has nearly 90 percent accuracy and it can be used in nearly all tracks across Pakistan.

The system, if used, can be a next step for the betterment of road safety conditions, planned movements, solution to over-saturated roads, analyzing driving behaviors and can lay emphasis on use of indicator lights on vehicles in Pakistan.

We can **integrate** *vehicle speed*, *driver seatbelt* and *number plate detection* system together. Driver's ***helmet detection system*** for *motor bikers* can also be introduced as a sub-feature. Use of ***Indicator lights*** and ***Behavior Analysis*** of drivers on roads can also be monitored by adding advanced features.

E- Chalan feature can be introduced for all the mentioned characteristics.

The system could be integrated with IP cameras and command and control center on ground in future.

It would not only improve road safety conditions but also strengthen security situation in our country.

15. REFERENCES

[1] S.Gulzar, F.Yahya, Z.Mir, R.Zafar, "Provincial Analysis of Traffic Accidents in Pakistan", SAVAP International, 2012

[2] Law Offices of Edgar Synder & Associates, "Seat Belt Statistics | How Many Injuries & Deaths Per Year", [ONLINE] Available at: https://www.edgarsnyder.com/car-accident/defective-products/seat-belts/seat-belts-statistics.html

[3] WHO, "Global Status Report on Road Safety: Time For Action", Switzerland 2009

[4] WHO, "Road Safety- Speed", 2004

[5] J.Xu, K.Song, "Study on Automatic Detection Method of Automobile Safety Belt Based on the Improvement of Adaboost Algorithm", 3rd International Conference on Mechatronics and Industrial Informatics, 2015

[6] H.Guo, H.Lin, S.Zhang, S.Li, "Image-based Seatbelt Detection", IEEE International Conference on Vehicular Electronics and Safety, July 2011

[7] B. Penchas, T.Bell and M.Monterio, "A Deep Learning Approach to Vehicle Speed Estimation", 2017

[8] Z.Lv, J.Xu, K.Zheng, H.Yin, P.Zhao and X.Zhou, "LC-RNN:A Deep Learning Model for Traffic Speed Prediction", IJCAI 2018

[9] W.Czajewski and M.Iwanowski, "Vision-Based Vehicle Speed Measurement Method", International Conference on Computer Vision and Graphics, Springer 2010

[10] S.Dogan, M.S.Temiz and S.Kulur, "Real Time Speed Estimation of Moving Vehicles from Side View Images from an Uncalibrated Video Camera", Sensors 2010

[11] A.Hussain, K.Shahzad, C.Tang, "Real Time Speed Estimation of Vehicles", World Academy of Science, Engineering and Technology, 2012

[12] M.A.Adnan, N.I.Zainuddin, N.Sulaiman and T.B.H.T.Besar, "Vehicle Speed Measurement Technique Using Various Speed Detection Instrumentation", BEIAC 2013

[13] L.G.C Wimalaratna and D.U.J Sunnandara, “Estimation of the Speeds of Moving Vehicles from Video Sequences”, Proceedings of the Technical Sessions, Institute of Physics – Sri Lanka, 2008

[14] D.Shukla, E.Patel, “Speed Determination of Moving Vehicles using Lucas Kanade Algorithm”, International Journal of Computer Applications Technology and Research, 2013

[15] T.Huang, “Traffic Speed Estimation from Surveillance Video Data”, IEEE Xplore CVPR, 2018

[18] W.Berger, “Haar Cascade”, 2018 [ONLINE] Available at: http://www.willberger.org/cascade-haar explained/#:~:text=Haar%20cascade,of%20Simple%20Features%22%20in%202001.

[19] A.Ryberg, “Camera Modelling and Calibration with Machine Vision Applications”, Ph.D Thesis, Sweden 2010

[20] A.Yilmaz, O.Javed and M.Shah, “Object Tracking: A Survey”, ACM Computing Surveys, Dec 2006

[21] R.C.Gonzalez, R.E.Woods, and S.L.Eddins, “Digital Image Processing using MATLAB”, CS 424 Lab, [ONLINE] Available at: ftp://www.cs.uregina.ca/pub/class/425-nova/Lab4/index.html

[22] S.Ahmed, W.Shahid, Lin, K.Walsh, A.U.Malik, “A Color Based Mango Maturity Device”, Master’s Thesis, NUST CEME, 2020

[23] N.K.Kanhere and S.T.Birchfield, “A Taxonomy and Analysis of Camera Calibration Methods for Traffic Monitoring Applications”, IEEE Transactions on Intelligent Transportation Systems, 2010

[24] A.Krizhevsky, I.Sutskever, G.E.Hinton, “Image Net Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems, 2012

[25] A.Karpathy, G.Toderici, S.Shetty, T.Leung, R.Sukthankar, L.F.Fei, “Large-scale Video Classification with Convolutional Neural Networks”, IEEE Xplore CVPR, 2014

[26] A.M.M. Ferreirinha, “Fully Automatic Roadside Video Camera Calibration and Traffic Analysis”, Master of Science in Electrical and Computer Engineering Thesis, University of Coimbra, Portugal, September 2016

[27] R.S.R.Dheekonda, S.K.Panda, M.D.N.Khan, M.Al-Hasan, S.Anwar, “Object Detection from a Vehicle using Deep Learning Network and Future Integration with Multi-Sensor Fusion Algorithm”, SAE International, 2017

[28] C. Szegedy, W.Liu, Y.Jia, P.Sermanet, S.Reed, D.Anguelov, D.Erhan, V.Vanhoucke, A.Rabinovich, “Going Deeper with Convolutions”, IEEE, CVPR, 2015

[29] K.Haugsdal, “Edge and Line Detection of Complicated and Blurred Objects”, Maters Thesis, Norwegian University of Science and Technology, June 2010

[30] B.Jung and G.S.Sukhatme, “Detecting Moving Objects using a Single Camera on a Mobile Robot in an Outdoor Environment”, 8th Conference on Intelligent Autonomous Systems, March 2004

[31] D.W.Wicaksono, B.Setiyono, “ Speed Estimation on Moving Vehicle Based on Digital Image Processing”, International Journal of Computing Science and Applied Mathematics, February 2017

[32] T.Lin, M.Marie, S.Belongi, L.Bourdev, R.Girshick, J.Hays, P.Perona, D.Ramanan, C.L.Zitnick and P.Dollar, “Microsoft COCO: Common Objects in Context”, Feb 2015

[33] O.Stenroos, “Object detection from Images using Convolutional Neural Networks”, Master’s Thesis, Aalto University, July 2017

[34] W.Wu, V.Koziksky, M.E.Hoover, R.P.Loce, “Vehicle Speed Estimation using a Monocular Camera”, SPIE 2015

[35] B.Makwana, P.K.Goel, “Moving Vehicle Detection and Speed Measurement in Video Sequence”, International Journal of Engineering Research & Technology (IJERT), Vol. 2, October 2013

[36] R.Tal, J.H.Elder, “An Accurate Method for Line Detection and Manhattan Frame Estimation”, York University 2012